

Découvertes de genres musicaux avec PySpark

• • •

Montréal Python 57 ~ 14 mars 2016

@deuxpi

La mission

Déterminer le style musical à partir de l'analyse des pièces écoutées.

Utiliser PySpark d'une façon originale.

Faire tout rouler sur un seul laptop.

Apprentissage automatique

aka Machine Learning

MIR : Music Information Retrieval

Extraction

Création d'un vecteur de « features » à partir d'une base de données musicale déjà étiquettée.

Modélisation

Création d'un modèle qui transforme les vecteurs et permet de faire une prédiction.

Classification

Grâce au modèle, on obtient de la nouvelle information qu'on peut analyser et tirer des conclusions.

Quelques exemples d'extractions

Énergie

L'intensité moyenne du son, la variation de cette intensité. La complexité du son.

Analyse spectrale

La distribution en fréquence du son, les hauteurs de notes qui reviennent plus souvent.

Rythme

La fréquence des événements musicaux (les notes) et on estime le tempo le plus probable.

Timbre

La texture du son, les instruments de musique utilisés.

Coffre à outils

- Jupyter
 - Python
 - librosa, numpy
 - Scikit-learn
 - Pandas
 - PySpark - Spark SQL
 - ffmpeg
-

Spark

Spark

- Moteur de calcul et de traitement de données
 - Construit pour du calcul distribué
 - Permet d'écrire des applications en Java, Scala, Python ou R
-
- Spark Core : style fonctionnel, replace bien Map Reduce
 - Spark SQL : sur-couche optimisée qui utilise SQL comme interface
 - Spark Streaming : traitement de données en continu
 - MLlib : machine learning
 - GraphX : analyse de graphes



Surprise ! Le BinaryType de Spark SQL ne fonctionne pas correctement sur Python 2.7



Python 3 ?

DEMO

Merci ! Questions ?

@deuxpi